

서술어/비서술어 명사 구조의 생산성*

김 동 성
(고려대학교)

Kim, Dong-Sung. 2012. Productivity of Predicative Nominal and Non-predicative Nominal Constructions. *Linguistic Research* 29(2), 403-434. In this study, we investigate productivity of nominal and non-nominal constructions in Korean. According to Lee (2005), nominal predicative construction is more morphologically productive than non-predicative nominal construction. We use the productivity index proposed in Baayen (1991, 1992, 1993, 2001). The result reveals that the productivity hypothesis in Lee (2005) does not hold the validity. Unlike Lee (2005), productivity of predicative/non- predicative nominal constructions is based on the argument structure, inducing consideration on the syntax-semantics (and/or syntactic structures). In order to capture such structure notion, we propose the type-token ratio of morphological tag concatenates which reflect syntactic and/or syntax-semantic information. Also, we find out that our proposal of the type-token ratio is more effective in explaining distribution of productivity. (Korea University)

Keywords Predicative Nominal and Non-predicative Nominal Constructions, Morphological Productivity, Type-Token Ratio, Statistical Models, Productivity of Construction, Syntax-Semantics Productivity

1. 서론

서술성 명사는 자신의 핵인 구에서 논항 부여가 가능하지만, 비서술성 명사는 논항 부여가 불가능하다. (1a)의 ‘연구’는 ‘철수’와 ‘수학’을 논항으로 취하는 서술어 명사이다. 반면에 ‘비빔밥’은 비서술성 명사로 논항을 취하지 못하므로, (1b)의 ‘전주’는 ‘비빔밥’의 수식어이지, 논항이 되지 못한다.

* 먼저 심사를 담당한 심사위원들께 감사를 드린다. 올바르게 명확한 지적이 있어서 논문의 형식과 내용이 향상될 수 있었다. 심사위원 중 한 분의 지적과 같이 한국어에서 서술어/비서술어 구분은 중요하다. 본 연구의 논지에 많은 도움을 줄 것이다. 그러나 이 논문에서는 서술어/비서술어 구분과 관련된 주장을 통한 선행적 이론의 토대를 구축하기 보다는 경험주의적 시각에서 논문을 저작하였으며 자료에 기반을 두고자 하였다. 논문에서 밝힌 바와 같이 서술어/비서술어 구조에 대한 논의들은 영킨 실타래와 같으며, 이 연구를 시작으로 도전하고자 한다. 마지막으로 논문의 모든 오류는 필자에 있음을 밝혀둔다.

- (1) a. 철수가 수학 연구를 했다.
 b. 전주의 비빔밥은 맛있다.

이선웅(2005)은 서술어/비서술어 명사 구조의 차이점을 다음과 같이 설명하고 있다. 비서술어 명사 구조는 합성어 구조로도 활용되거나 합성어가 되기 때문에 형태론적 과정을 겪어서 어휘부에 등재되지만, 서술어 명사 구조는 형태론적 과정보다 의미-통사 구조에 의존하기 때문에 논항 부여, 논항 실현과 같은 구조화 과정을 겪게 된다는 것이다. 이를 바탕으로 이선웅(2005)은 형태론적 과정을 겪는 비서술성 명사 구조보다 구조화 과정을 겪는 서술성 명사 구조가 더 생산적이라는 가설을 제시하였다. 특히 ‘어절1(보충어 성격의 명사/대명사 포함) + 어절2(서술어/비서술어 명사 포함)’의 경우에 서술어/비서술어 구조 간에 형태론적인 생산성에 있어서 차이가 난다는 가설을 제시하였다(이선웅 2005: 147).

이선웅(2005)의 이와 같은 문제 제기는 우선적으로 형태론적 생산성의 문제와 연관되어 있다. 그러나 서술어/비서술어 구조는 비단 형태론적 생산성의 문제점이 아닌 통사, 의미-통사와 같은 여러 언어학적 측면에 걸쳐있다. 따라서 이선웅(2005)의 가설은 영킨 신타래와 같은 많은 언어학적 문제점들을 포괄하고 있다. 여러 문제들 중 본 논문에서 다음의 질문에 답변하고자 한다.

- (2) a. 서술어/비서술어 명사 구조는 형태론적인 생산성의 차이를 보이는가?
 b. 만약 (2a)에서 제기한 형태론적 생산성의 차이가 서술어/비서술어 구분과 연관이 없다면 무엇이 생산성의 차이를 결정하는가?
 c. 생산성이 구조화 과정과 같은 의미-통사적 측면의 정보에 기인한다면 어떠한 방식으로 측정해야 하는가?

본 연구에서는 코퍼스를 활용해서 생산성을 측정한다. 형태론적 생산성의 문제에 대해서 Baayen (1992, 1993)의 생산성 척도 측정 방식을 활용한다. 연구에서는 (2b)와 같이 형태론적 생산성의 차이가 서술어/비서술어 구분과 연관성이 없는 것으로 나타났는데, 어떠한 자질이 생산성과 연관이 있는지 논의하게 될 것이다. 또한 (2c)의 문제에 대해서 본 연구에서는 통사적 생산성은 의미-통사적 행태인 논항 구조에 기인한다는 Barddal (2008)의 제안에 따라 코퍼스를 통해서 의미-통사의 틀을 살펴볼 수 있는 기재인 타입-토큰 비율(type-token ratio)을 제안한다. 이 분석 기재는 의미-통사 구조를 파악하는 단서를 제공하지만 Baayen (1992, 1993)이 제안한 형태론적 생산성 척도와도 상관성이 높고, 어휘 분포의 모형을 더 정확히 예측한다. 그러므로, 본 논문에서 제기한 타입-토큰 비율은 (2b, c)에서 제기한

문제인 형태론적 생산성의 문제와 구조화 과정의 문제 모두를 해결할 수 있는 유용한 기재이다.

논문의 구성은 다음과 같다. 2절은 생산성에 대한 선행 연구로 언어 이론적 틀에 대한 연구와 통계적 연구를 살펴본다. 경험적 연구인 통계적 연구가 이론적 틀과 어떻게 연결되어 있는지를 알기 위해서 두 측면을 논의한다. 3절은 연구 대상, 실험 내용, 실험 결과에 대한 분석이다. 실험은 의미 분석이 된 21세기 세종계획 형태의미 분석 말뭉치 750만 어절[이하 ‘의미분석코퍼스’]을 활용하였다. 실험은 형태론적 분석과 의미-통사 구조 분석을 위해서 타입-토큰 비율을 활용하였다. 실험에 대한 결과 분석으로 통계적으로 의미-통사를 설명하기 위한 모형인 타입-토큰 비율이 형태론적 생산성 척도와 연관성이 높은 것을 설명하였다. 4절은 이 논문의 결론이다.

2. 생산성에 대한 선행 연구

2.1 생산성과 언어학

상대적으로 한정된 어휘가 무한한 언어 사용에 활용되기 때문에 생산성은 형태론적으로 중요하다. 굴절과 파생은 여러 측면에서 다르게 나타나는데, 생산성은 굴절과 파생을 구분하는 요소 중 하나이다. 굴절은 생산성이 높는데 반해서 파생은 굴절에 비해서 높지 않다. 이와 같이 생산성은 형태론 이론을 설명하는 구성 요소 중 하나이다. 생산성의 많은 연구들이 형태론에서 활발히 논의되었는데, 여러 연구 중 Aronoff (1976)은 생산적인 단어 형성만이 형태론적 연구의 대상이라고 규정하였다. 더 나아가서 Aronoff (1976)는 어휘부와 관련해서 생산적인 단어 생성 규칙들로 만들어진 어휘들은 규칙으로 다루어야 하며 어휘부에 기재할 필요성이 없다고 주장하였다. Aronoff (1976)은 생산성을 다음과 같이 설명한다.

“... we count up the number of words which we feel could occur as the output of a given WFR [word-formation-rules] (which we can do by counting the number of possible bases for the rules), count up the number of actually occurring words formed by that rule, take a ratio of the two, and compare this with the same ratio for another WFR. In fact, by this method we could arrive at a simple index of productivity for every WFR: the ratio of possible to actual words (1976: 36).”

Aronoff (1976)이 제시한 생산성의 개념은 단어 형성 규칙의 실제 대상이 되는 어휘들의 수와 원칙으로 적용이 가능한 어휘들의 수의 비율이다. 또한 어휘적 관계성이 생산성에 직접적인 영향을 준다고 주장한다. Aronoff (1976: 45)에서는 생산성이 의미적 일관성이 연관이 있으며, 어휘부의 기재와 생산성이 상반되는 관계가 있음을 주장하였다.

전체적인 문법 모델에서 생산성은 어휘부의 형상과 문법의 흐름을 포착하는데 중요한 역할을 한다. 단어 형성은 생산성과 직접적인 연관을 갖고 있는데, 생산적인 부류의 단어 형성은 규칙적인 어휘부가 체계 일관성(system congruity)을 갖고 있다는 것을 의미한다(Wurzel 1984). Anderson (1982: 594)은 통사부가 음운부로 발견되는 표면 구조로서의 문법 모델을 제시하고 이 틀에서 체계화된 생산적인 어휘부를 주장하였다. 반면에 Di Sciullo and Williams (1987)은 통사부가 음운부의 경계를 설정하는데, 생산성이 어휘부에 기재되는 단어들의 목록을 결정하지 않고, 통사부의 형성과 연관성이 있다고 주장하였다.

한국어의 경우에 시정곤(1999)와 최형용(2000)은 생산성의 문제가 한국어 형태론에서 중요한 위치를 차지하며, 어휘부의 기재에 있어서 고려해야 할 대상이라고 주장하였다. 또한 시정곤(2006)은 생산성의 개념을 단어 형성 과정, 더 확장하면 형태론적 과정에 모두 적용되는 것으로 규정하였다. 생산성을 공시의 관점에만 적용할 것과 생산성의 정도가 높은 경우를 규칙화 차원에서 논의할 것을 주장하였다.

구조적으로 생산성의 개념은 언어학에서 형태론적 연구에 집중되었다. 그러나, 생산성은 언어 전반에 걸쳐서 취급되어야 할 중요한 요소이며, 특히 형태론적 측면이 통사적, 의미적 관점과 연결되어 있는 문제에 있어서는 생산성은 비단 형태론적 과정의 문제만이 아니다. 서술어/비서술어 명사의 구분을 논의한 강범모(2001)의 논의를 살펴보면, 형태론적 측면의 문제는 통사론적 측면의 문제에서 기인하는 것으로 보인다. 이러한 틀에서 생산성 논의가 시작되면 생산성은 형태론 단독의 측면에서 논의되어야 할 요소로는 여겨지지는 않는다.¹ 또한 서술어/비서술어 명사 구조를 다룬 이선웅(2005: 136-146, 166-168)은 서술어, 비서술어 구조의 의미-통사적 측면의 논항 구조에서 기인한 구조가 형태론적으로 재구조화 되는 과정을 설명하고 있다. 따라서, 생산성 논의가 비단 형태론적 측면이 아닌 통사구조나 의미-통사 구조로 확장되어야 할 필요성이 제기 된다.

¹ 강범모(2001: 17)에서 서술어 명사 구조 [[X + 하] + 계]의 구조는 통사론적 규칙에서 기인하는데, 형태론적으로 재구조화(reconstruction)의 과정을 통해서 [X + [하 + 계]]의 구조로 변한다고 설명한다. 이러한 점은 통사적 측면이 형태적 측면과 연결되어 있는 것으로 판단할 수 있으며, 생산성 논의가 형태와 통사에 단독으로 적용되는 것이 아니고 형태와 통사의 두 측면에 걸쳐서 적용되어야 하는 것으로 확대할 수 있다.

통사론적 생산성을 다룬 Barddal (2008)은 형태론적 생산성 개념을 통사론적 생산성으로 확장해야 한다고 주장한다. 통사적 생산성은 통사적 창조성과 같으며, 언어 사용자가 제한된 규칙으로 무한한 수의 언어 사용을 만들어내는 것과 같다. 구체적으로 단어들의 유형인 구절 구조가 반복적으로 재귀되는 통사 현상을 구절 구조 규칙이 포착할 수 있다(Chomsky 1957). 또한 언어 이론적으로 언어 능력이란 언어의 규칙을 이해하는 것과 같다는 신념으로도 확장될 수 있다(Chomsky 1965). 더 나아가서 Barddal (2008: 29)은 통사적 생산성은 얼마나 많은 사용을 어떠한 규칙이 예측할 수 있는가의 개념으로도 귀결될 수 있으며, 토큰(언어 사용의 예)과 타입(예측 규칙) 구분과 상호작용한다고 주장한다.

Barddal (2008)은 통사적 생산성은 의미적 부분인 논항 구조에 종속적이라고 주장한다. 동사나 서술어의 논항 구조에 따라 구조가 생성되므로, 더 많은 다른 유형의 타입을 가진 동사나 서술어는 더 높은 타입의 빈도를 의미하고, 타입-토큰 대비는 의미-통사적 관점의 생산성을 나타낸다(Barddal 2008: 29).

이러한 Barddal (2008)의 주장은 이전의 의미, 의미-통사와 연관된 이론들과 연관되어 있다. 생산성과 연관된 이론들을 정리하면 다음과 같다. 의미-통사적인 측면에서 Jackendoff (1990)는 개념 구조(Conceptual Structure) 이론을 제시하였는데, 통사 분석된 구절 구조에 의미 구조가 정보 구조로 해석되게 해서 의미-통사의 생산성을 포착하고자 하였다.² Fillmore (1977)는 의미, 의미-통사의 관점에서 언어 사용의 생산성은 프레임(Frame)이라는 구조화된 지식의 표현을 통한 프레임 의미론(Frame Semantics)을 제시하였다. Minsky (1974)가 제시한 지식 도해의 기법중 하나인 프레임을 활용한 Fillmore (1977)은 구조에 대한 해석이 의미에 대한 틀에서 기인한다고 주장한다. 통사적 생산성과 의미적 생산성이 접합점이라는 것은 Levin (1993)의 경험적 연구에서 나타난다. Levin (1993)이 경험적 자료에 근거해서 동사의 교체 현상인 통사적 행태가 의미적 행태와 동일함을 제시하였는데, Levin and Rappaport (2005)는 의미역의 논항 실현(argument realization)이 사건 개념화(event conceptualization)로 나타나는 것이 통사-의미의 접합점이라고 주장하였다.³ 의미적 관점에서 Pustejovsky (1995)는 단어의 의미가 맥락에 맞게 변화하는 다의성도 구조적으로 포착될 수 있다고 주장하였는데, 의미의 다형적(polymorphic) 생산성도 언어학의 큰 틀 안에서 이론화될 수 있다.^{4,5}

2 Jackendoff(1990)은 Katz(1972)의 성분 분석(Componential Analysis)을 발전시켜서 해석한 것이다.

3 언어 습득의 관점에서 Conwell and Demuth(2007)은 통사적 생산성을 교체 현상에서 연구하였다.

4 Pustejovsky(1998)은 의미론 측면에서 생산성의 관점을 포착하고자 하였는데, 통사 형식의 생산성과 의미 행태를 논의하였다.

5 이 이외에 의미론적 관점에서 여러 연구들이 발견된다. 그 중 철학적인 틀에서 Davidson(1980)은 동작(action)이나 움직임(motion)도 하나의 사건 구조의 틀 안에서 분석된다고 주장하였는데,

여러 이론들을 간략하게 종합하면 의미, 의미-통사와 연관된 여러 이론들에 따르면 의미의 생산성은 문장 구성 성분인 논항들의 의미 역할에 달려 있다. 의미역이라 불리는 행위주격(Agent), 상대격(Patient), 처소격(Locative), 목적격(Goal), 대상격(Theme) 등등이 어떻게 분포하고, 주어, 동사, 목적어와 같은 문장 요소로 나타나는가가 생산성과 연관되어서 중요하게 취급되어야 할 부분이다. 바꾸어 말하면, 통사 행태는 의미적 행동에서 비롯되므로 이를 생산성의 관점에서 파악하기 위한 방식이 요구된다. 또한 본 연구의 논의점인 서술어/비서술어 명사 구조의 생산성은 형태론적 대상과 연결되어 있으므로, 이를 통합적으로 살펴볼 수 있는 관점이 필요하다.

2.2 코퍼스를 활용한 생산성 연산

전반적으로 생산성은 형태론적 단어 형성과 국한되지 않고, 통사, 통사-의미, 의미와 같은 여러 분야에 걸쳐서 나타난다. 따라서 복합적인 틀에서 생산성과 연관된 현상을 분석할 필요성이 있다. 경험적 자료에 근거한 연구는 대용량의 코퍼스에 기반을 두는데, 파생 어미의 생산성과 같은 형태론에 기반을 둔 Baayen (1992, 1993), Baayen and Liber (1991)의 연구와 논항 실현과 교체 현상과 연관된 통계 현상을 다룬 Gries and Stefanowitsch (2003, 2004), Bresnan et al. (2007)의 의미-통사 연구가 발견된다.

코퍼스를 활용하는 것은 통계적인 행태를 관찰하기 위해서이다. 예를 들어서, N개의 토큰을 가진 코퍼스에서 특정 어휘 V(N)개의 타입이 관찰되는 경우를 고려해 보자. 이때 일정한 크기의 토큰으로 코퍼스를 분할하고 어휘의 토큰이 어떻게 커지는지를 살펴보면 어휘 성장 곡선(vocabulary growth curve)을 관찰할 수 있다.

Parsons(1990)는 이를 토대로 문장의 의미를 사건 의미론(Event Semantics)이라는 모델을 통해서 포착하고자 하였다.

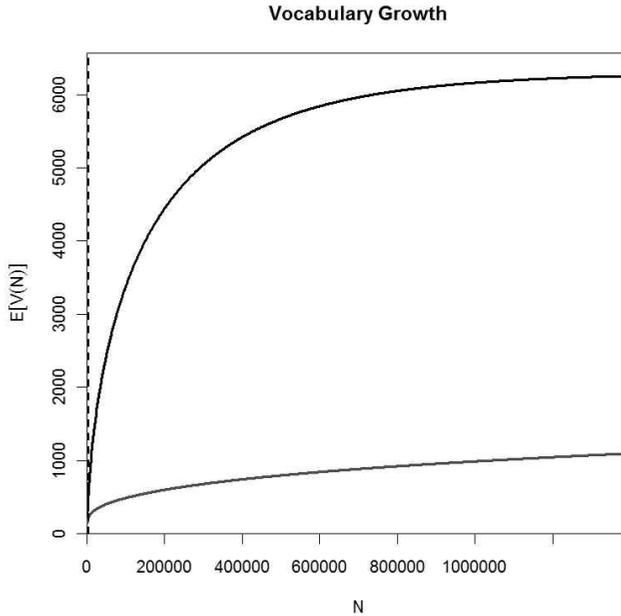


그림 1. 어휘 성장 곡선

그림 1은 200,000 어휘 토큰들 마다 얼마나 많은 어휘 토큰들이 있는가를 측정하고 곡선을 그린 것이다. 그림 1에서 위에 위치한 곡선은 아래에 위치한 곡선보다 생산성이 더 높다. 생산성이 높은 어휘는 성장이 빠르고 급격한 어휘 성장 속도를 보이다가 완만한 곡선으로 발전하는데 비해서, 생산성이 낮은 어휘는 성장을 거의 찾아보기 힘든 정도의 성장 속도를 보인다.

어휘 성장 속도를 측정하기 위한 것이 Baayen (1992, 1993), Baayen and Liber (1991)에서 제시한 통계 모델이다. 이 모델은 Aronoff (1976)에서 제시한 생산성을 수치적으로 해석한 것인데, 기본적으로 V가 어휘 타입이고, S가 단어 형성 규칙 WFR이 적용 가능한 경우일 때, 생산성 척도 I는 (3)과 같이 관찰된다.

$$(3) I = \frac{V}{S}$$

공식 (3)은 이상적인 수치로 단어 형성 규칙 WFR이 적용 가능한 잠재적 대상들을 측정해야 하는데 사실상 불가능하다. 따라서 코퍼스에서 발견되는 통계적으로 유의미한 수치로 전환해야 하는데, Baayen (1992, 1993)이 제시한 생산성 척도는 P로 (4)와 같이 계산되는데, 전체 토큰 N과 빈도가 1인 타입 개체의 수인 n_1 의

비율로 측정한다.

$$(4) P = \frac{n_1}{N}$$

빈도가 1인 어휘들을 hapax legomena라고 하는데, 이 어휘들이 해당 어휘와 얼마나 많이 공기하는지를 측정하고 이를 전체 코퍼스 토큰과 비율적으로 나타낸다. 이 측정법은 Baayen (1992, 1993)에 따르면, 어휘 성장 곡선의 기울기를 표현한다. 그림 2와 같이 전체 어휘 토큰 N 과 개별 어휘 타입 $V(N)$ 의 비율은 Δ 로 (4)의 생산성 척도 P 와 동일하다. 다르게 말하면 Δ 는 곡선의 기울기로 성장 곡선이 어느 정도 상향하는가의 정도를 나타낸다. 측정지수 P 는 여러 통계 모형 중 지프(Zipf) 모형, Yule-Simon 모형, Waring-Herdan-Muller 모형과 같은 빈도와 빈도 순서의 분포를 나타낸다(Baayen and Liber 1991).⁶

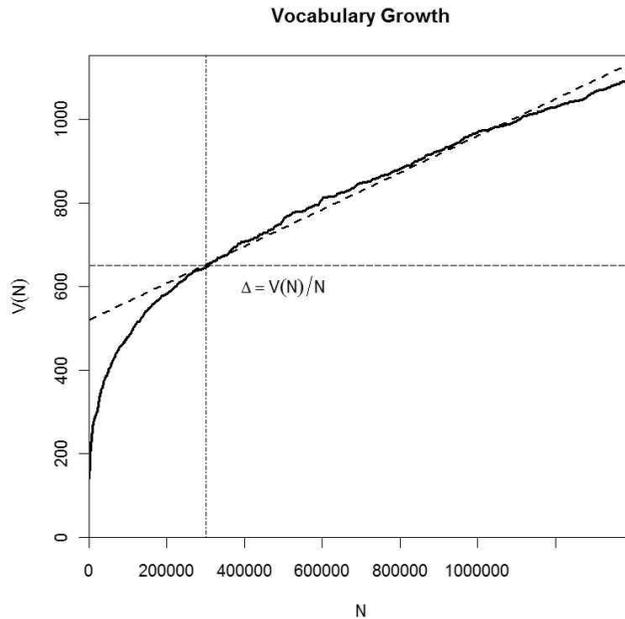


그림 2. 어휘 성장 곡선의 기울기

한국어의 경우에 이광호(2006)은 Baayen (1992, 1993)의 측정법을 수정해서 과

⁶ 이 모형들은 지수 분포 모형으로 일반적인 정규 분포 모형과 다르다. 자세한 논의는 Baayen(2001)을 참조.

생 접미사의 생산성을 측정할 것을 주장하였다. 이와 다르게 강범모·차준경(1996)과 시정곤(2006)은 Baayen (1992, 1993)의 측정법이 생산성에 고려될 수 없다고 주장하며, 단순한 통계량의 측정일 뿐 구조에 대한 접근이 아니라는 것이다. 이들은 생산성 척도 P 가 높은 경우에 빈도가 낮은 어휘가 포함되어 있고, 생산성 척도 P 가 낮은 경우에도 빈도가 높은 경우도 있어서 척도가 생산성을 대변하지 못한다고 주장한다.

그러나, Baayen (1992, 1993)의 논의를 자세히 살펴보면, 생산성 척도는 구조적으로 생산성이 높은 경우와 낮은 경우를 설명한다. 이를 위해서 전역(global) 생산성 지수인 P^* 를 제시하였는데, 생산성 척도 P 와 어휘 유형 개수인 V 를 비율로 나타낸다. 이 측정 방식을 활용하면 단어 형성 규칙인 WFR의 통계량을 측정하고, 개별 어휘의 생산성을 측정할 수 있다고 한다.

P^* 가 높은 경우는 생산성이 높다고 할 수 있는데, 두 가지 경우에 가능하다. 첫 번째 경우에 생산성 척도 P 가 높고 유형 개수 V 가 적은 경우 P^* 는 높게 측정될 수 있고, 두 번째 경우에 생산성 척도 P 도 높고 유형 개수 V 도 상대적으로 높으면 P^* 가 상대적으로 높게 측정된다. 두 번째 경우는 문제가 되지 않으나 첫 번째 경우에 유형 개수 V 가 적기 때문에 상대적으로 적은 수의 개체가 코퍼스에 발견되기 때문에 문제를 제기할 수 있다. 강범모·차준경(1996)과 시정곤(2006)은 빈도에 근거한 수치 측정에서는 낮은 빈도로 산출되기 때문에 전체 개수가 적은데 어떻게 생산성 척도 P 가 높은가를 문제로 제시한다.

이에 대해서 Baayen (1992, 1993)에서는 생산성 척도 P 는 코퍼스에 상대적으로 측정되는 것이 아닌 코퍼스에 종속적으로 측정된다고 주장한다. 특히 Baayen (1991, 1992)은 어휘 크기에 따라서 생산성 척도 P 가 달라지는 것을 포착하였다. 더 자세히 설명하면 두 개의 다른 코퍼스 상에서 생산성 척도 P 가 다르게 측정되기 때문에 특정 코퍼스에서 측정된 P 가 어휘 크기에 따라서 달라진다(Baayen 1991). 생산성 척도 P 는 활용하는 코퍼스 상에서 측정되는 수치이므로, 코퍼스에 종속되어 있다. 또한 빈도라는 통계 수치도 코퍼스에 종속되어 있다. 빈도가 코퍼스의 전체량에 대비해서 측정되는 것이기 때문에, 코퍼스의 크기를 바꾸어서 측정하면 빈도 또한 다르게 측정된다. 따라서 생산성 척도 P 와 빈도에 의존적으로 생산성 측정하는 것은 코퍼스에 독립적인 측정이라고 할 수 없다. 또한 빈도가 낮고 높은 것도 하나의 코퍼스 상에서의 측정법일 뿐 전체 언어 사용을 대변하는 통계량 지수는 아니다.

반면에 생산성 지수 P^* 는 생산성 척도 P 와 어휘 유형 개수인 V 를 비율로 측정되기 때문에 일정한 양의 코퍼스들에서 상대적으로 균일하게 측정된다. 따라서 전역 생산성 지수 P^* 의 통계는 전체 생산성 측정에 있어서 코퍼스에 독립적으로 측

정되기 때문에 생산성 측정에 유용한 지표이다.

그렇다면 문제는 생산성 척도 P 는 생산성 측정에 있어서 유용한지에 대한 의문점이 제기된다. Baroni and Evert (2006)은 Baayen (1992, 1993)의 생산성 척도 P 에 근거해서 어휘 생산성을 분석하였는데, 이 연구에서 생산성의 측정은 어휘 성장 곡선과 연관성이 높다고 주장하였다. 또한 Baayen (1991, 1992, 1993, 2001)의 일련의 연구들에서 발견되는 생산성은 어휘 성장과도 연관성이 있다. 따라서 생산성 지수 P^* 는 여러 코퍼스를 아우르는 거시적인(macro) 측면에서의 전역 생산성 측정법으로 활용되고, 생산성 척도 P 는 어휘 성장과 같은 미시적인(micro) 측면에서의 생산성 측정법으로 활용된다면 유용성에 대한 해결이 될 것이다.

Gries and Stefanowitsch (2003, 2004)와 Bresnan et al. (2007)의 의미-통사 연구는 논항 구조의 논항 실현에 대한 연구로 교체 현상의 생산성에 대한 연구이다. 이전 절에서 논의한 바와 같이 동사가 가진 의미에 따라서 개별 논항이 실현되는데, 논항 실현은 각기 다른 의미에 따라서 각기 다른 구조로 표현된다. 프레임 의미론, 개념 구조, 교체 현상 등이 이러한 의미-통사의 논항 실현에 대한 언어 이론적 연구들이다.

교체 논항의 예를 하나만 들어 보자. 동사 ‘cover’는 (6)과 같이 교체가 불가능하지만, ‘spray’는 (5)와 같이 논항 교체가 가능하다.

- (5) a. The farmer sprayed water on the plants.
- b. The farmer sprayed the plants with water.
- (6) a. Monica covered the baby with a blanket.
- b. *Monica covered a blanket over a baby.

이 경우에 ‘spray’는 with-대상격이 가능하지만, ‘cover’는 over-대상격이 불가능하다. Gries and Stefanowitsch (2003, 2004)는 동사의 부류를 이와 같은 구조에 따른 확률에 근거한 통계량으로 동사를 분류하였다. 결과적으로 의미가 유사성이 높은 동사들끼리 서로 통계적으로 동일한 집합으로 분리되어서 의미-통사가 서로 접합점을 보이는 것을 경험적으로 증명하였다. 예를 들어서 (5)와 (6)의 경우에 ‘spray’와 ‘cover’는 각기 다른 의미 집합으로 구분되게 된다.

Bresnan et al. (2007)는 처소격이나 대상격으로 실현되는 두 개의 서로 다른 논항 구조의 경우에 어떠한 것이 처소격이나 대상격 논항으로 실현되는데 기여하는지를 통계적으로 연구하였다. 여러 가지 언어학적 요인을 통계적 변인으로 취급하였는데, 예를 들면 대상격의 animacy 유무, 대상격의 성, 수, 인칭 등등과 같은 구체적인 언어학적 대상들이 요인들이 된다.

Gries and Stefanowitsch (2003, 2004)와 Bresnan et al. (2007)는 교체 현상에서 논항 실현을 연구의 대상으로 하였다는 점에서 유사성이 있지만, 생산성의 관점은 크게 다르다. Gries and Stefanowitsch (2003, 2004)는 문장의 구조에 (예를 들어서 with-처소격이나, over-대상격인지에 따른 구조) 따라서 생산성이 결정되는 반면에 Bresnan et al. (2007)은 논항의 animacy 유무와 같은 언어학적 요인에 따라서 생산성이 결정된다고 주장하였다. 그러나, 두 연구 모두 생산성이 구조와 연관되어 있다는 점에서 Baayen (1992, 1993)의 형태론적 생산성 연구와 다르다. 다시 말하면, Baayen (1992, 1993)은 형태론적인 문제에 국한되어서 통계량에 의한 생산성을 주장한 반면에 두 연구는 생산성을 문장 구조에 기반을 두고 있다고 주장하였다.

언어학적 측면에서 문장 구조의 생산성은 Fillmore (1977)의 프레임 의미론과 Jackendoff (1990)의 개념 구조 이론 등의 구조적 틀을 활용한 생산성에 가깝다. 구조적 틀에 대한 연구는 직접적으로 통사적 행태가 의미적 행태와 동일하다는 Levin (1993)과 Levin and Rappaport (2005)의 통사-의미 접합점에 대한 경험적 연구를 코퍼스를 활용한 통계적 연구로 발전시킨 것이다. 흥미로운 점은 Gries and Stefanowitsch (2004)의 연구에서는 코퍼스에서 발견되는 문장의 구조들의 집합체인 연구조(collostructure)의 군집들을 통계적으로 도출해내는데, 이 연구조의 군집들은 Levin (1993)이 경험적으로 통사 의미를 분류한 것과 유사성이 높다는 것을 주장한다. Gries and Stefanowitsch (2003, 2004)와 Bresnan et al. (2007)의 의미-통사 구조의 생산성에 대한 연구는 논항 구조의 생산성이 구조에 의한 생산성과 동일하다는 것을 보여준다.

본 연구에서 논의할 서술어/비서술어 명사 구조는 형태론적 문제와 연관성이 있지만, 논항 구조 및 논항 실현 등과 같은 의미-통사 문제가 관련되어 있으므로 구조적 문제와도 연관성이 있다. 연구에서는 이러한 점을 고려해서 통계적 측면에서 형태론적 측면과 의미-통사적 측면을 모두 고려할 수 있어야 한다는 점을 고려하였다. 따라서 Baayen (1991, 1992, 2001)의 일련의 연구에서 제시된 형태론적 생산성을 고려하면서, 동시에 의미-통사에서 비롯된 구조적 생산성도 고려할 수 있는 방법을 고려하였다. 이선웅(2005)의 가설이 형태론적 가설에 기초하고 있으므로 형태론적 관점을 고려하였다. 또한 서술어/비서술어 명사 구조는 구조적 생산성과 연관성이 있으므로 구조적 생산성도 고려되어야 한다. 연구에서 발견한 것은 형태론적 유형들이 논항 구조와 연관성이 높다는 것이다. 3절과 4절에서 이러한 부분을 자세히 논의할 것이다.

3. 이론과 코퍼스를 활용한 생산성에 대한 접근

3.1 서술어/비서술어 구조와 생산성⁷

서술어 명사는 명사구에서 핵이 되며 논항을 취할 수 있다. (7b, c)의 ‘조사’라는 명사는 동사로 활용된 (7a) ‘조사하다’의 논항 구조와 동일하다.

- (7) a. 검찰이 참고인을 조사하다.
 b. 검찰의 참고인 조사
 c. 검찰 참고인 조사

반면에 비서술어 명사는 논항을 취할 수 없고, ‘수식어 + 비서술어’ 구조만 가능하다. (8b, c)의 경우에 (7b, c)와 다르게 논항 구조가 불가능하다.

- (8) a. 철수가 찹쌀로 밥을 하다.
 b. *철수의 찹쌀 밥⁸
 c. *철수 찹쌀 밥
 d. 철수의 찹쌀밥

명사는 크게 서술어와 비서술어 명사로 구분된다.⁹ 여러 연구들이 서술어/비서

7 서술어/비서술어 명사 구분은 문장의 이해와 연관되어서 논항 구조의 판별에 대한 기초를 제공하므로, 언어학적으로 중요하다. 지금까지 여러 연구들이 서술어/비서술어에 대한 구분에 대한 연구를 제시하였는데, 그 중 강범모(2000, 2001)는 여러 통사적 조건들을 제시하였다. 또한 홍재성(1998)은 서술어 명사의 경우에 서술어 명사는 논항 구조와 같은 어휘 특질을 담당하고, 문장의 상, 시제 등의 문법적 역할을 하는 ‘지지동사(support verb)’가 사용되는 것에 기초해서 서술어 명사들을 분류하였다.

8 여기서 ‘찹쌀밥’은 합성어 구조이어야 한다. 따라서 가능한 구조는 ‘수식어 + 합성어’만 가능하다. 의미적으로는 (8b)나 (8c)가 ‘철수가 찹쌀로 밥을 하다’의 의미나 ‘철수가 밥을 하지는 않았지만 찹쌀로 된 밥을 갖고 있다’ 중 하나를 가리킬 수 있다.

9 이미 앞서 논의한 바와 같이 서술어/비서술어 명사 구분에 대한 통사적 논의가 강범모(2000, 2001)에서 이루어졌다. 그런데, 이후 연구인 이선웅(2005), 이병규(2009)의 논의와 같이 강범모(2000, 2001)이 제시한 기준이 강범모(2000, 2001)에서 언급한 바와 같이 ‘폐 쓸만한 다소 불완전한’ 대상들이다. 따라서 어떠한 기준들이 정확하게 판별 조건이 되는지에 대한 부분은 명확하지는 않다. 그러나, 이후 박현아-강범모(2006)의 연구에서 논의한 바와 같이 판별 조건들은 정도성에 근거하였다고도 할 수 있다. 더 자세히 설명하면, 일정한 조건에 의해서 가부가 결정되는 것이 아니고, 여러 다면체적인 조건들이 얽혀져서 증감률(gradient)에 의해서 결정된다는 것이다. 본 연구는 생산성과 연관된 부분에 초점이 맞추어져 있으므로, 서술어/비서술어 명사 구분이나 서술어 명사 분류 등의 문제를 논의의 대상으로 하지는 않는다. 그러나, 논의의 명확성을 위해서 서술어/비서술어 명사의 구분이 되는 논항 구조, 논항 부여, 상과 같은 통사적 행태 등을 거론한다. 서술어/비서술어 명사 구분

술어 명사 구조에 대한 논의를 제시하고 있지만, 그 중 홍재성(1999), 강범모(2001), 이선웅(2005), 이병규(2009)의 연구에서 생산성에 대한 논의를 살펴보면 다음과 같다. 서술어/비서술어 명사 구조의 생산성은 이선웅(2005)에서 다음과 같이 직접적으로 논의되고 있다.

“...왜 ‘보충어+비술어명사’ 구성이 ‘보충어+술어명사’보다 합성어로 등재되는 일이 많은지 명확히 밝히기는 매우 어려워 보이지만, 형태론에서 말하는 ‘생산성’의 개념과 관련지을 수 있을 듯하다. ‘보충어+술어명사’의 결합은 특정한 의미적 조건만 충족되면 거의 무제한적으로 만들어낼 수 있으나, ‘보충어+비술어명사’의 경우는 그렇지 않다 (2005: 147).”

이선웅(2005)의 논의를 자세히 살펴보면 다음과 같다. 서술어 명사 구조는 비서술어 명사 구조에 비해서 생산성이 높으므로, 생산성이 서술어/비서술어 명사 구조의 구분에 직접적으로 사용될 수 있다는 것이다. 또한 형태론적 생산성의 개념이 직접적으로 적용될 수 있다는 것이다. 이 가설은 여러 논쟁의 소지가 많다. 먼저, 확인되어야 할 점은 과연 서술어 명사 구조가 비서술어 명사 구조에 비해서 생산성이 높은지에 대한 점이다.

또한 이선웅(2005)의 가설에서 많은 문제점이 발견되는데, 그 중 하나는 서술어/비서술어 명사 구조는 구조적 생산성인데 어떻게 형태론적 생산성과 연관이 있는가이다. 이전 2.1절에서 논의한 바와 같이 서술어/비서술어 명사 구조에서 통사론적 규칙인 구조적 규칙들이 재구조화와 같은 과정을 통해서 형태론적 과정을 겪는 것이 강범모(2001), 이선웅(2005)에서 설명되고 있다. 따라서, 통사적 행태가 형태론적으로 연결되어 있으므로, 형태론적 생산성이 통사나 의미-통사의 구조적 생산성과 연결되어 있다고 주장될 수 있다.¹⁰

그러나, 문제가 되는 것은 (7, 8)과 같이 서술어 명사만이 서술어 성격의 명사로 서술어 성격의 행태를 보인다는 것이다. 이러한 논증은 술어 명사 구조가 논항 부여, 논항 실현, 상(aspect)과 같은 서술어 성격과 연관성이 있다는 이병규(2009)의 주장을 뒷받침한다. 서술어/비서술어 명사 구분이 상적 표현인 ‘중’의 가능 여부를 통해서 판별이 가능하다는 강범모(2001), 이병규(2009)의 주장으로도 발전할 수 있다.¹¹

은 중요한 문제로 향후 연구되어야할 주제로 미룬다.

¹⁰ 또한 강범모(2001)은 ‘-음’ 파생형이 서술어/비서술어 구분에 연관성이 있다는 지적을 하였다. 이를 고려하면, 형태론적 과정이 구조적 과정과 직접 연관이 있다는 주장이 가능하다.

¹¹ 그러나, 비서술어 명사로 분류된 ‘지진’의 경우에 ‘중’과 결합이 가능하다. 이 경우에 상적 표현으로 활용되어서, ‘지진이 발생한 중’의 의미 해석이 가능하다. 필자가 인터넷 검색을 한 결과 이러한 사

- (9) a. 빨래 중, 빨래하는 중
 b. *밥 중, 밥하는 중

또한 홍재성(1999)의 논의와 같이 서술어 명사의 경우에 (9a)의 ‘일으키다’와 같은 ‘기능동사’와 연관된 구조도 가능하다.¹²

- (10) a. 코리올레이너스가 로마를 향해 반란을 일으켰다.
 b. 코리올레이너스의 로마를 향한 반란

(10a)와 (10b)가 모두 올바른 형태인 것은 서술어 명사 ‘반란’이 서술어 성격을 띠게 되어서 논항 구조와 연관된 문법성이 발현되었기 때문이다. 반면에 비서술어 명사는 서술 동사의 사용이 필수적이다.

- (11) a. 서울시가 시청 앞에서 음악회를 열었다.
 b. *서울시의 시청 앞에서 음악회

따라서, 서술어 명사 구조의 경우에만 논항 구조와 연관된 생산성이 술어적 성격의 문법성을 나타내기 위해서 발현된다고 주장할 수도 있다. 2.1절에서 논의한 바와 같이 논항 구조는 의미-통사적 측면에서 생산성과 연관이 높다.

그러나, 이전 연구의 논의를 잘 살펴보면 비서술어 명사 중 관계명사인 경우에는 논항 구조의 성립이 가능하다(강범모 2001). 예를 들어서 (12)의 ‘어머니, 아버지’와 같은 친족 명사인 경우에는 논항으로 ‘철수, 영희’가 사용된다.

- (12) a. 철수의 어머니
 b. 영희의 아버지

또한 강범모(2001)는 ‘팔, 다리, 얼굴’과 같은 신체부위나 ‘가지, 마개, 지붕’ 등의 부분어를 나타내는 비서술어 명사도 논항 구조 설정이 가능하다고 주장한다. 강범모(2001)의 주장을 확장하면 대부분의 비서술어 명사의 경우에 논항 구조가 가능하다.

용이 발견된다. 예를 들어서 ‘일본 대지진 중에 천당과 지옥을 오고간 쓰나미 강아지 사진들...’과 같은 용례가 발견된다.

¹² ‘기능동사’는 시제, 상, 법, 양태와 연관된 문법성과 연관되어 있다(홍재성 1999, 강범모 2001: 5)

- (13) a. 앞집 사람의 개
 b. 언어학의 교수
 c. 대한민국의 관례
 d. 전주의 비빔밥

(7, 8)에서 제기한 바와 같이 논항 구조의 여부가 서술어/비서술어 명사 관별에 중요하지만, 비서술어의 경우에 논항 구조가 완전히 불가능하지 않다. (12, 13)과 같이 제한된 수준에서 비서술어 구조의 경우에 논항 구조가 가능하다. 따라서 의미-통사 측면에서 논항 구조와 연관된 생산성은 서술어나 비서술어 명사 모두 적용되어야 한다.

그러므로, 생산성과 연관된 측면에서 확인되어야 할 점은 다음과 같다. 먼저 이선웅(2005)의 가설에 따라 형태론적 생산성의 차이가 서술어/비서술어 구조에서 나타나는지 살펴보게 될 것이다. 다음 절에서 형태론적 생산성의 문제를 의미코퍼스를 통해서 논의하게 될 것이다. 이 문제 이외에도 다음과 같은 여러 문제들을 고려해야 한다. 이선웅(2005)의 가설과 같이 형태론적 생산성의 차이가 서술어/비서술어 구조적 생산성을 설명할 수 있는가? 만약 이선웅(2005)의 가설이 잘못 되었다면 무엇이 생산성의 차이와 연관성이 높은가? 의미-통사와 연관되어서 논항 구조와 연관된 생산성을 어떻게 포착할 수 있는가?

다음 절에서 언어 자료를 통해서 생산성을 논의하게 될 것이다. 형태론적 측면, 형태-통사적 측면, 의미-통사적 측면 모두를 설명할 수 있는 방법이 무엇인지 논의할 것이다. 이를 위해서 더 포괄적으로 확장된 시도가 필요하므로, 본 논문에서는 언어 자료의 범위를 ‘어절1(보충어 포함) + 어절2(서술어/비서술어 명사 포함)’로 확장하였다. 2.2절에서는 언어 자료에서 이러한 의미-통사적 측면의 생산성을 포착하기 위한 여러 연구들을 소개하였는데, 어떠한 방식이 의미-통사적 측면의 생산성을 포착할 수 있는지도 다음절에서 논의할 것이다.

3.2 언어 자료를 활용한 실험 및 결과

서술어/비서술어 명사 목록은 연구자에 따라 약간씩 차이가 있다. 강범모(2000)는 (14), 이선웅(2005)은 (15)와 같은 목록을 제시하였다.

- (14) a. 서술어 명사
 공부, 노래, 도움, 믿음, 반란, 부, 부상, 불, 불공, 빨래, 산책, 소문, 손해, 수모, 숙제, 실의, 암살, 애착, 암전, 연구, 연민, 오류, 운동, 원

한, 자유, 조용, 체포, 파괴, 폭격, 행복, 호통, 환멸, 휘방

b. 비서술어 명사

감기, 개, 결, 고양이, 나무, 눈, 떡, 머리, 밖, 밭, 밥, 병, 비, 손, 아버지, 안, 어머니, 얼굴, 음악회, 잠, 정신, 종교, 지진, 책상, 춤, 틀니, 팔, 편지

(15) a. 서술어 명사

간주, 감격, 감동, 개발, 걱정, 건강, 건설, 결정, 결혼, 경영, 계획, 고려, 고통, 공개, 공격, 관계, 관심, 구타, 국교, 귀향, 귀환, 긴장, 남획, 달성, 대표, 대화, 도피, 모색, 반란, 반응, 배신, 번역, 부정, 비난, 비판, 상대, 생각, 성공, 성과, 수렴, 수여, 수입, 신중, 안심, 애교, 약속, 여행, 연구, 연유, 오해, 운영, 유래, 자유, 잔치, 잘못, 재회, 정리, 조작, 질문, 집념, 집회, 철퇴, 출발, 충격, 치부, 타격, 탐닉, 통합, 파견, 파괴, 파문, 행복, 행사, 헌신, 혐오, 확인

b. 비서술어 명사

가지, 감기, 개념, 계기, 고기, 공, 관례, 교수, 나라, 나무, 냄새, 눈, 능력, 단위, 도구, 도사, 매체, 문헌, 물건, 바다, 바닥, 박사, 밖, 밭, 밥, 방향, 밭, 번개, 병, 비, 뿌리, 사례, 사진, 색깔, 선수, 성격, 세대, 손, 시간, 실력, 아버지, 안, 어머니, 얼굴, 용어, 원인, 잎, 재료, 저서, 정신, 지진, 책, 천둥, 천지, 친분, 혈안

실험에서 사용된 의미분석코퍼스는 형태 분석이 된 코퍼스에 의미 분석을 첨가한 것으로, 국립국어원에서 발간한 표준국어대사전의 표제어 단위에 표제어의 의미를 분류한 어깨번호까지 포함하였다. 예를 들어서, 그림 3과 같이 ‘외국, 기업, 외면’은 표준국어대사전에 등재되어 있는 표제어들인데, 해당 표제어들은 여러 의미를 갖고 있어서 각각 02, 01, 03과 같은 어깨번호로 구분된다.

외국	외국__02/NNG
기업들이	기업__01/NNG + 들/XSN + 이/JKS
외면한다	외면__03/NNG + 하/XSV + ㄴ다/EC

그림 3. 의미분석코퍼스의 예

(14, 15)의 목록을 대상으로 의미분석코퍼스에서 해당 어깨번호까지 일치하는 어휘들을 연구의 대상으로 하였다. 선정된 목록들이 포함된 2어절 중에서 이선웅

(2005)의 가설이 포함된 ‘어절1(보충어 포함) + 어절2(서술어/비서술어 명사 포함)’ 구조에 해당하는 어절 연쇄를 추출하였다. 첫 번째 앞 어절에는 명사, 대명사류가 포함되게 하였고, 두 번째 뒤 어절에는 대상 어휘가 포함되게 하였다.

이와 같은 목록들을 대상으로 통계/의미적으로 관련성이 높은 표준국어대사전에서 다의성 목록¹³, 의미분석코퍼스에서 발견되는 어휘 빈도, 명사포함 어절의 빈도, hapax legomena가 포함된 어절의 빈도를 추출하였다. 이를 토대로 분석표를 표 1과 같이 작성하였다.

표 1. 비서술어 이선웅(2005)의 추출의 예

어휘	표준국어대사전 어깨번호	어휘 빈도	다의성 목록	명사 포함 어절 빈도	hapax legomena	<i>P</i>
관례	관례__03	138	0	38	1	0.0263
도사	도사__10	64	3	17	1	0.0588
물건	물건	1772	5	463	1	0.0021
...

추출된 자료를 대상으로 (4)의 생산성 척도 *P*를 연산하였다. 생산성과 연관된 질문에 대해서 첫 번째로 이선웅(2005)의 가설인 서술어 명사 구조가 비서술어 명사 구조보다 형태론적 생산성이 높다는 것에 대한 부분을 입증하였다. 강범모(2000)과 이선웅(2005)의 목록들을 대상으로 서술어/비서술어 전체 목록의 생산성 척도 *P*를 연산하였다. 결과적으로 두 연구에서 제시한 서술어/비서술어 목록의 생산성 정도인 생산성 척도 *P*는 다르게 연산되었다.

표 2. 목록 전체의 생산성 척도 *P*

	서술어 <i>P</i>	비서술어 <i>P</i>
강범모(2000)	0.009417	0.006731
이선웅(2005)	0.007668	0.012617

강범모(2000)에서 제시한 목록 중 서술어 명사 구조의 생산성 척도 *P*가 더 높지만, 반대로 이선웅(2005)에서는 비서술어 명사 구조가 생산성이 더 높다. 이러한

¹³ 표준국어대사전의 어깨번호는 동철동어어나 동음동의어를 구분한 단위로 표기되어 있다. 의미분석 코퍼스는 표준국어대사전의 어깨번호까지만을 구분하였기 때문에, 하위 다의는 구분하지 않았다. 예를 들어서 ‘물건’의 경우에 표준국어대사전에는 5가지 다의가 구분되어 있는데, ‘[1] 일정한 형체를 갖춘 모든 물질적 대상, [2] 제법 어떠한 구실을 하는 존재, ...’ 등등과 같이 구분되어 있다.

표 2의 결과는 이선웅(2005) 자체의 가설과도 위배된다. 또한 두 연구에서 제시한 목록의 생산성 지표가 다르게 산출되었다는 것은 서술어/비서술어 구분이 생산성 척도와 연관성이 없어 보인다. 다르게 말하면, 서술어/비서술어 구분이 생산성에 미치는 영향보다 다른 요인이 생산성에 미치는 영향이 더 크다고도 말할 수 있다.

생산성 척도 P 는 형태론적 생산성으로 서술/비서술 구분과 연관성은 없다. 따라서 어떠한 요인이 생산성과 직접적으로 관련이 높은지 연산하기 위해서 표 1에서 제시한 어휘 빈도, 다의성 목록, 명사 포함 어절 빈도, hapax legomena와 생산성 척도 P 가 어떠한 연관성이 있는지 통계적으로 포착하고자 하였다. 통계적 관련성은 두 변인들 간의 선형적 관계성을 측정하는 상관 분석으로 포착할 수 있다.¹⁴ 연구에서 활용한 것은 피어슨 상관 계수(Pearson correlation coefficient)로 두 변인들 간의 표준 편차를 공분산(covariance)으로 나누는데, 두 변인이 상관성이 없는 통계량과 상관성이 있는 통계량의 비율로 측정한다.¹⁵ 측정된 결과는 표 3과 같다.

표 3. 상관성 분석 1

	구분	어휘 빈도	다의성 목록	명사 포함 어절 빈도	hapax legomena
강범모 (2000)	서술	0.16275	0.25541	0.12962	0.40886
	비서술	-0.15899	-0.1363	-0.1564	-0.00711
이선웅 (2005)	서술	-0.07931	-0.09945	-0.09065	0.0478
	비서술	-0.11007	-0.07814	-0.03926	0.66509

여러 변인들이 생산성 척도 P 와 상관성의 정도가 일관성이 없다. 표 3에서 생산성 척도 P 와 변인들 간에 측정되는 상관성 중 hapax legomena가 상관성 정도가 가장 높다. 생산성 척도 P 연산이 hapax legomena와 전체 어휘 빈도와와의 비율로 측정되기 때문에 생산성 척도 P 는 hapax legomena와 가장 유사성이 높다. 그러나, 전반적인 수치는 어휘 빈도, 다의성 목록, 명사 포함 어절 빈도, hapax legomena는 생산성 척도 P 와 일관된 상관성을 보이지 않는다. 결론적으로 어휘 특징을 나타내

¹⁴ 간략하게 상관 분석을 소개하면 다음과 같다. 상관성은 두 개의 다른 요인이나 변인들 간에 측정되는 것으로 상관성이 높다는 것은 선형적으로 두 변인들이 밀접하게 관련되어 있다는 것을 말한다. 또한 상관성의 정도는 양의 상관성과 음의 상관성, 무상관성으로 나눌 수 있다. 양의 상관성은 하나의 변인의 증가가 다른 변인의 증가로 나타나며, 음의 상관성은 반대로 감소로 나타나고, 무상관성은 두 변인들은 아무 상관성이 없다는 것이다.

¹⁵ 측정 수치는 +1과 -1 사이에서 산출되는데, +1은 완전한 양의 상관성, -1은 완전한 음의 상관성, 0은 무상관성을 의미한다. 수치가 높으면 높을수록 상관성의 정도가 높아지는데, 양의 상관성의 경우에 0~0.1은 무상관성, 0.1~0.3은 미미한 상관성, 0.3~0.5는 약한 상관성, 0.5~0.7은 더 강한 상관성, 0.7~1.0은 가장 강한 상관성을 의미한다.

는 요인들인 어휘 빈도, 다의성 목록, 명사 포함 어절 빈도, hapax legomena도 생산성 척도 *P*와 상관성이 거의 없다.

이와 같이 형태론적 측면의 생산성과 관련된 여러 요인들이 아무런 연관성이 없다는 것 이외에 또 다른 측면에 대한 고려도 필요하다. 2절과 3.1절에서 논의한 바와 같이 서술성/비서술성 명사 구조의 생산성은 형태론적 생산성 이외에 통사, 의미-통사측면의 생산성과 연관되어 있다. 이러한 측면을 고려하기 위해서 ‘어절 1(보충어 포함) + 어절2(서술어/비서술어 명사 포함)’ 구조 중 ‘어절1’의 형태론적 결합 구조를 추출하였다.

1. 진행상태를	진행__02/NNG + 상태__01/NNG + 를/JKO	조사하는
2. 대한교련이	대한교련/NNP + 이/JKS	조사한
3. 교육청에서	교육청/NNG + 에서/JKB	조사를

그림 4. 형태론적 결합 구조 추출의 예

그림 4에서 ‘어절2’의 ‘조사’가 포함된 명사 구조에 대해서 ‘어절1’의 경우에 ‘NNG NNG JKO’, ‘NNP JKS’, ‘NNG JKB’와 같이 다른 형태론적 결합 구조가 추출된다. 이와 같은 형태론적 결합 구조는 통사, 의미-통사 측면을 반영한다. 예를 들어서 서술어 명사 ‘조사’의 경우에 세종전자사전을 찾아보면 주절의 동사가 ‘받다, 되다’로 바뀌면 서술어 구조가 다르게 나타난다.

(16) a. 받다

A씨가 혐의에 대해 검찰로부터 조사를 받다

b. 되다

그 일이 한 시민에 의해 조사가 되다

세종전자사전의 ‘조사되다’와 ‘조사받다’ 용언의 논항 구조를 살펴보면 (16a, b)는 각각 (17a, b)의 논항 구조로 나타나고, 각각의 논항 구조에 따라 다른 격표지를 지닌다.

(17) a. X = A씨, Y = 혐의, Z = 검찰

X = Goal, Y = Theme, Z = Agent

X = N-이/가, Y = N-에 대해, Z = N-를

b. X = 일, Y = 한 시민

$X = \text{Theme}, Y = \text{Agent}$

$X = \text{N-이/가}, Y = \text{N-에 의해}$

(16)에서 ‘조사’를 중심으로 이전 어절의 형태론적 정보를 추출하면 각각 ‘NNG JKB, VV EC’와 같다. 각각의 동사에서 논항에게 부여되는 격표지나 의미역도 각각 다르게 나타나며, 이러한 형태론적 정보는 논항 구조 정보가 된다. 따라서 ‘NNG JKB’와 ‘VV EC’는 각각 ‘조사를 받다’, ‘조사가 되다’와 연관된 논항 구조 정보이다. 이와 같은 이론적 근거에 바탕을 두고 형태론적 결합 구조를 추출하고, 각각의 구조 타입과 구조 토큰을 대상으로 타입-토큰 비율을 측정하였다. 이를 토대로 표 4와 같은 타입-토큰 비율을 산출하였다.

표 4. 비서술어 이선웅(2005)의 타입/토큰 비율 산출의 예

어휘	표준국어대사전 어깨번호	구조 타입	구조 토큰	타입-토큰 비율
관례	관례_03	17	38	0.44737
도사	도사_10	7	17	0.41176
물건	물건	56	463	0.12095
...

표 4의 타입/토큰 비율을 생산성 척도 P 와 피어슨 상관 계수 측정을 통한 상관성을 측정하였다. 여기서 통계적 유의성을 위해서 t -test를 시행하였는데, 모두 유의성 검증을 통과하였다.

표 5. 상관성 분석 2

	구분	타입/토큰 비율	t	df	p-value
강범모(2000)	서술	0.65579	3.7865	19	0.001248
	비서술	0.817016	6.6459	22	1.108e-06
이선웅(2005)	서술	0.895819	14.2532	50	2.2e-16
	비서술	0.284522	1.9462	43	0.05819

이선웅(2005)의 비서술 명사 유형들의 상관성 정도가 적은 경우만을 제외하고 모두 상관성의 정도가 높은 것으로 나타났다. 따라서 생산성 척도 P 는 형태론적 결합 구조와 상관성이 있으며, 구조에 대한 생산성을 나타내는 타입-토큰 비율과 상관성이 높다. 다시 말하면, 생산성 척도 P 는 어휘 빈도나 다의성 등과 같이 어휘

차체의 요인에 따른 생산성의 정도와 상관성이 없고, 구조적 생산성의 정도와 상관성이 있다고 해석된다.

3.3 구조의 생산성

타입-토큰 비율은 하나의 타입이 얼마나 많이 사용되었는가를 나타낸다. 어휘의 타입-토큰 비율에서 비율이 높은 경우는 더 많은 수의 타입이 사용된 경우로, 다양하게 다른 어휘의 유형들이 활용되었다고 할 수 있다. 소설과 신문 기사를 비교하면 소설의 비율이 신문 기사보다 더 높다. 소설은 미사나 수식어구가 많으며, 다양한 등장인물이 등장하므로 다양한 어휘들이 사용되므로 비율이 높다. 반면에 신문 기사와 같은 성격의 텍스트는 적은 수의 분명한 표현법을 가진 어휘들을 사용해야 하므로 비율이 낮다. 마찬가지로 구조에 대한 타입-토큰 비율이 높으면 여러 구조가 다양하게 사용되었으며, 반대로 낮으면 일정한 구조가 반복적으로 사용된 것이다. 언어학적으로 Fillmore (1977)의 프레임 의미론이나 Jackendoff (1990)의 개념 구조의 틀 안에서 비율이 높으면 여러 다양한 의미-통사 유형이 사용되었다고 말할 수 있다. 또한 Levin (1993), Levin and Rappaport (2005)의 논항 실현과 교체 현상을 통해서도 여러 논항 구조가 다양하게 실현되었다고도 할 수 있다.

이선웅(2005)나 강범모(2001)의 연구에서 형태론적 과정이 통사적 과정과 연관되어 있다는 주장은 형태론적 생산성과 통사적 생산성 사이에 연관성이 있다는 것으로도 해석된다. 따라서, 연구에서 활용한 통사적 생산성을 나타내는 형태론적 연쇄의 타입-토큰 비율이 형태론적 생산성을 나타내는 hapax legomena와 연관성이 높다는 사실이 발견되면 통사적 생산성이 형태론적 생산성과 직접적인 연관성이 있다.¹⁶

16 심사자 중 한 분이 형태론적 생산성과 통사적 생산성 사이의 연관성이 무엇인지 질문을 하면서, ‘형태론적 구조는 통사적 규칙이 참조할 수 없는 불투명한(opaque) 구조’라는 주장을 하였다. 이 견해에 따르면, 형태론적 과정이 통사적 과정과 연관성이 없다. 형태론적 과정이 통사적 과정과 별개로 처리된다면 실질적으로 형태론적 과정과 통사적 과정은 아무런 연관성도 없을 것이다. 그러나 필자는 형태론적 구조는 통사적 규칙을 참조해야 하는 투명한(transparent) 구조라고 생각한다. 그러나 반대로 통사적 규칙이 형태론적 규칙에 직접적인 연관이 있다면, 통사적 규칙으로부터 형태론적 규칙이 도출이 된다. 따라서, 강범모(2001)의 [[X + 하] + 게]가 [X + [하 + 게]]로 재구조화되는 과정을 일련의 연쇄적 과정으로 도출해 낼 수 있다. 강범모(2001)의 설명과 같이 재구조화라는 것은 구조를 다시 해석하는 것으로 ‘불투명한 구조’에서는 재구조화가 불필요하고, 형태론적으로 새로운 규칙만 생성되면 된다. 다시 말하면 형태론적 규칙이 통사적 규칙을 참조하지 않고 새로운 규칙으로 구조를 생성하면 된다. 따라서 강범모(2001)에서 주장한 재구조화는 개념적으로 필요가 없을 것이다. 필자는 형태론적 과정이 통사적 과정과 연관성이 있고, 형태론적 구조가 통사적 측면의 규칙을 참조해야 한다고 주장한다. 예를 들어서, 접사(clitic)는 통사적 측면의 규칙이지만, 형태론적 측면의 생성이다. 그러므로 심사자의 의견과 달리 필자는 형태론적 과정이 통사적 측면과 밀접하게 연결되어 있다고 판단한다.

통계적으로 텍스트의 특성을 잘 나타내는 것이 빈도가 1인 어휘들인 hapax legomena이다. Baayen (1992, 1993)에서는 이 어휘들이 형태론적 규칙 WFR의 특성을 잘 나타낸다고 주장하는데, 본 연구에서는 hapax legomena을 활용하면 규칙들의 타입들에 대한 통계적 수치가 분석할 수 있다. Hapax legomena가 발견되는 타입들은 발생 빈도가 높은 형태론적 연쇄이다. 모든 hapax legomena가 발생하는 형태론적 유형을 조사해보면 통계학적으로 4분위수(quartile) 중 가장 높은 75~100%에 해당하는 4분위수에서만 발생한다. 그림 5에서 회색으로 음영이 있는 부분으로 정규 분포 모형을 4등분할 경우에 수치가 가장 높은 분위이다.

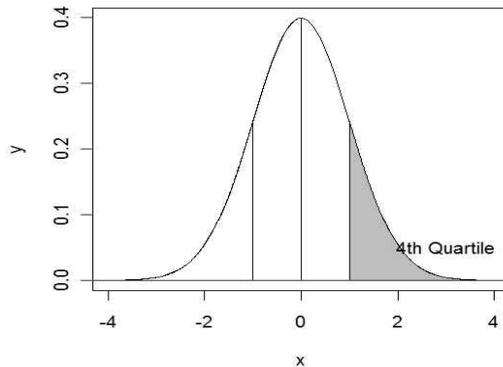


그림 5. 4분위수 모형

예를 들어서 서술어 명사 ‘자유’의 경우에 1,957개의 토큰이 발견되며, 그 중 148개의 타입이 발생하므로 타입-토큰의 비율은 0.00756이다. 이 중 hapax legomena가 나타나는 유형은 4개인데, 4분위 수의 경계는 수치적으로 5이고, 5이상이면 4분위 수에 해당한다. 해당 각 타입의 토큰 수는 표 6과 같다.

표 6. ‘자유’의 hapax legomena 타입-토큰

유형	토큰 수
NNG NNG	24
NNP	25
NNG JKB	163
NNG JKG	413

이러한 hapax legomena의 행동 방식은 다음과 같이 해석될 수 있다. 빈도가 낮은 어휘는 언어 사용이 매우 적은 행동 방식을 갖게 되며, 이러한 행동 방식으로

인해서 선택하는 구조는 가장 익숙한 구조에 활용한다. 다시 말하면, 가장 사용 빈도가 높은 구조를 활용해서 빈도가 낮은 어휘를 활용하게 된다. 따라서 hapax legomena는 구조적으로 의미-통사 구조와 연관성이 높다.

전체적으로 구조와 생산성 척도 P 를 같이 살펴보면 더 흥미로운 결과를 발견할 수 있다. 강범모(2000), 이선웅(2005)에서 제시한 서술어/비서술어 목록들에 대한 legomena (어휘 빈도)에 해당하는 생산성 척도 P 와 구조에 대한 타입-토큰 비율의 피어슨 상관 계수를 측정해 보면 표 7과 같이 legomena에 따라 상관성이 다르게 나타난다. 같은 legomena에서는 동일한 상관성이 나타났다. Hapax legomena와 dis legomena (빈도 2), tris legomena (빈도 3)이상 deca legomena (빈도 10)이하, deca legomena 초과를 살펴보면 상관성이 1에 가까운 0.9 이상을 보이고 있다. 따라서 균일한 수준의 빈도 내에서는 상관성 척도 P 와 구조에 대한 타입-토큰 비율의 상관성이 높은 것으로 나타난다.

표 7. 빈도 분포에 따른 피어슨 상관 계수

어휘 빈도	강범모(2000)		이선웅(2005)	
	서술어	비서술어	서술어	비서술어
1	0.9102	1	0.97629	0.84772
2	0.92747	0.98762	0.95972	0.93202
3~10	0.98504	0.94447	0.89385	0.92478
10초과	0.36584	0.86555	0.76915	0.35444

이와 같이 형태론적 생산성을 나타내는 hapax legomena는 구조적 생산성과 직접적으로 연관되어 있다. 표 6의 예에서와 같이 구조적으로 생산성이 높은 구조에 형태론적 생산성이 높게 나타난다. 이것은 생산성이 구조와 밀접하게 연관되어 있으며, 형태론적 과정이 통사적 과정과 연관되어 있다고도 할 수 있다. 또한 표 7과 같이 구조적 생산성은 legomena와 일치한다. 그러므로, 생산성의 개념을 형태론적 과정과 의미, 통사적 과정이 서로 연결되어 있는 것으로도 확장할 수 있다. 이러한 관점은 Baayen (1992)에서 제시한 형태론적 관점의 생산성이 구조적 관점의 생산성으로 다시 해석되어야 하는 것을 보여주며, 더 나아가서 구조에 대한 생산성이 형태, 의미, 통사의 여러 문법 단계에 걸쳐 있는 것으로도 확장해서 해석할 수 있다.

형태론적 생산성이 통사적 제약에 따라 결정되기도 한다. 영어의 경우에 시제와 상과 같은 굴절 접미사는 동사에만 국한되고, 명사형 접미사 '-ness'는 형용사, 접두사 're'는 대부분이 동사에만 가능하다. 형태론적으로 형태론적 과정을 제한하는 요소는 형태론에만 기인하지 않는다. 통사, 의미, 음운, 어원 등의 다양한 제약

조건이 형태론의 생산성을 제한한다. 결론적으로 생산성의 개념은 형태론적 과정에만 국한되어있지 않고, 다양한 측면의 과정과 연계되어 있다. 더 나아가서 형태론적 과정이 통사적 과정과 연계되어 있기 때문에, 형태론적 생산성을 통사적 생산성과 구분하는 것도 어렵다.

구조의 타입-토큰 비율은 Baayen (1992, 1993)에서 제시한 생산성 척도 P 에 비해서 생산성 분포를 왜곡하지 않는다. 2.2절에서 논의한 Baayen (1992, 1993)의 생산성 척도인 P 와 어휘 유형 개수인 V 의 비율인 전역 생산성 지수인 P^* 를 활용하면 이 문제에 정확하게 접근할 수 있다. 생산성 척도인 P 와 어휘 유형 개수 V 를 대변하는 빈도에 대한 산포도 그래프를 그려보면 그림 6과 같다. 여기서 그림 6은 이선웅(2005)의 서술성 명사의 생산성 척도 P 와 어휘 유형 개수인 빈도와 의 산포도 그래프이다.

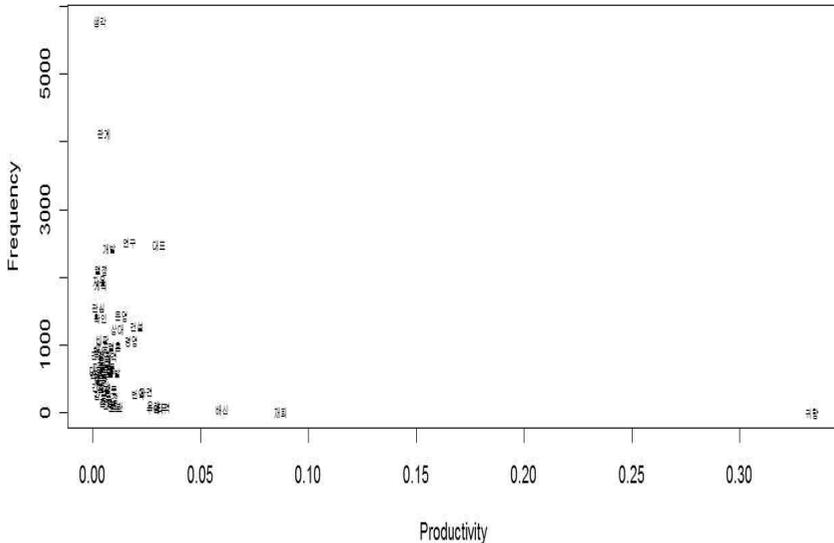


그림 6. 생산성 척도 P 와 빈도

전역 생산성 지수 P^* 는 빈도와 생산성 간의 관계성을 나타낸다. 산포도에서 빈도는 높지만 생산성 척도가 낮은 개체는 y축에 가깝게 나타나고, 생산성 척도는 높지만 빈도는 낮은 개체는 x축에 가깝게 나타난다. 여기서 대부분의 어휘들이 y축, x축에 가깝게 분포하는 것은 빈도와 빈도순을 나타내는 빈도 스펙트럼 곡선과 유사하다. 그림 7은 이와 같은 형태의 곡선의 예인데, 빈도가 높은 것의 타입은 적고, 빈도가 낮거나 1인 것의 타입은 매우 높다. 그림 7은 『Alice in Wonderland』

소설의 단어 빈도순과 빈도의 그래프이다. 이러한 곡선은 지프 곡선이고 분포는 지프 분포로 알려져 있다.¹⁷

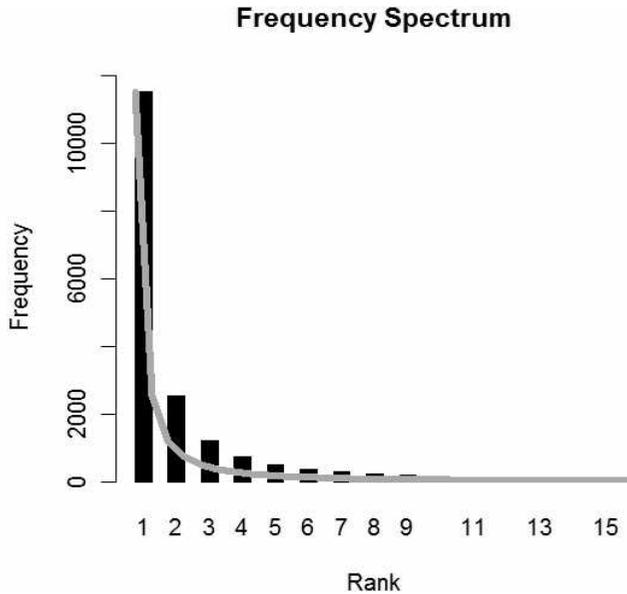


그림 7. 빈도 스펙트럼 곡선

Baayen (1993)에 따르면, 생산성 척도 P 는 범주형 통계량이지만, 전역 생산성 지수 P^* 는 빈도 수치 1에 근거한 생산성의 분포를 보인다고 한다. 또한 Baayen (2001)에서는 전역 생산성 지수 P^* 는 지프 분포에 근거한 지수형 분포 모형으로 그림 7의 분포를 따른다. Baayen (1991, 1992, 1993, 2001)에서 제시한 P^* 가 지수 모형이 되는 근거는 다음과 같다. 먼저 $N \rightarrow \infty$ 이 될 경우인 코퍼스의 어휘 수가 무한대로 커지면 $\frac{n_1}{N}$ 은 0이 된다. 다시 말하면, 코퍼스는 제한된 언어 사용을 수집한

17 간단히 지프의 법칙과 연관된 단어 분포를 설명하면 다음과 같다. 지프의 법칙은 단어의 빈도는 빈도순에 반비례한다는 것이다. 하나의 단어 w 에 대해서 단어의 빈도가 $f(w)$, 단어의 빈도순이 $r(w)$ 이면 $f(w) = \frac{C}{r(w)^a}$ 의 관계가 성립하는데, 여기서 $f(w)$ 는 가능한 데이터의 수준에 따라서 상수 C 와 a 에 의해서 결정된다. 이 관계는 지수형 분포로 관계를 선형적 관계로 바꾸어 관찰하려면 로그 선형성(log linearity)을 적용해서, $\log f(w) = \log C - a \log r(w)$ 로 바꾼다. 지프의 법칙에서 고빈도어가 지나치게 높고, 저빈도어가 지나치게 낮기 때문에 경험적으로 관찰하기 어려운 점이 있어서 일정한 인수 b 를 분모의 $r(w)$ 에 더해서 값을 조정하면 지프의 법칙은 $f(w) = \frac{C}{(r(w)+b)^a}$ 와 같이 바뀌게 된다. 이것은 지프-만델브로트의 법칙(Zipf-Mandelbrot's law)로 알려져 있다(Baayen 2001).

유한한 자료인데, 자료를 언어 사용과 동일하게 무한하게 수집하면 (18)과 같은 수학적 모델링이 가능하다.

$$(18) \lim_{N \rightarrow \infty} \frac{n_1}{N} = 0$$

따라서, 어휘가 무한대가 되면 생산성 척도 P 는 0으로 x 축과 동일하게 된다. 이론적으로 생산성 척도 P 는 그림 2와 같이 코퍼스에서 주어진 지점에서의 기율기와 같으므로, 주어진 hapax legomena 예측치 E 와 시간 t 라는 지점에서 (19)와 같이 구할 수 있다.

$$(19) P_{E,t+1} = \frac{\Pr(X(t+1)|\mu_{x(t+1)})}{\Pr(X(t+1)) \cap \Pr(\mu_{x(t+1)})} \\ = \frac{\Pr(X(t+1)) \cap \Pr(\mu_{x(t+1)})}{\Pr(\mu_{x(t+1)})} \\ \approx \frac{n_{1,E,t}}{N_{E,t}}$$

여기서 $N_{E,t}$ 가 ∞ 으로 수렴할지라도 $n_{1,E,t}$ 나 $N_{E,t}$ 는 실질적으로 0이 될 수 없다. 왜냐면, 무한한 코퍼스 상에서 한 지점인 $n_{1,E,t}$ 나 $N_{E,t}$ 는 모두 0이 아니며 대응되는 값이 존재한다. 따라서, 전체 생산성 척도 P 도 0이 될 수 없다. 결과적으로 (18)의 공식은 이론적으로 존재하지 실질적으로 존재하지 않는다. 결론적으로 현실적으로 (18)이 아닌 (20) 공식이 성립한다.¹⁸

$$(20) \lim_{N \rightarrow \infty} \frac{n_1}{N} > 0$$

18 여기서 (20)의 공식은 실제로 $\lim_{N \rightarrow \infty} \frac{n_1^N}{V^N} > 0$ 으로 바뀌어야 한다. Baayen (1992)에 따르면, 어휘의 숫자 V^N 가 증가하면 할수록 hapax legomena의 숫자인 n_1^N 도 무시하지 못할 숫자로 증가하게 된다고 한다. 다시 말하면 어휘 총 개수인 N 의 증가에서 의미 있는 hapax legomena의 증가는 어휘 유형에 비례해서 증가하게 된다는 것이다. 증가함수의 정의역과 치역은 (20)과 같이 $N \rightarrow \infty$ 으로 표시되는데, 생산성이 높은 어휘와 더불어 생산성이 낮은 어휘 모두가 사용되기 때문에 전체 어휘 개수인 N 이 정의역이어야 한다. 또한 Baayen (1992)은 생산성이 높은 경우에 어휘증가 함수는 $\lim_{N \rightarrow \infty} \frac{n_1^N}{V^N} > 0$ 를 반드시 적용해야 한다고 주장한다. 또한 Baayen (1992)은 실질적인 어휘 크기는 $\lim_{N \rightarrow \infty} V^N$ 으로 V^N 은 지프 분포에 따라 분포하는 제타 분포(Zetta distribution)를 따른다고 주장한다.

결과적으로 생산성 척도 P 는 0이 되지 않지만 0보다 크며 0에 수렴하기 때문에 x축을 따라 그려지는 지수 곡선 형태의 분포가 된다. 또한 생산성 척도 P 를 활용한 전역 생산성 지수인 P^* 도 생산성 척도 P 와 같이 지수 곡선 형태가 되며 그림 7과 같이 지프 곡선의 형태로 분포하게 된다(Baayen 1992, 1993). Baroni and Evert (2006)는 이탈리아어 단어의 생산성을 분석한 결과 생산성 척도 P 가 지수 형태의 분포를 보인다고 한다.

그림 6에 추세 곡선(trend curve line)과 같이 x축 값과 y축 값의 서로 일치하는 부분을 지나는 선을 그려 보면 그림 8과 같이 나타난다. 이 선을 따라가면 빈도 스펙트럼 곡선과 유사성이 있지만 일정 부분이 왜곡되어 있다.

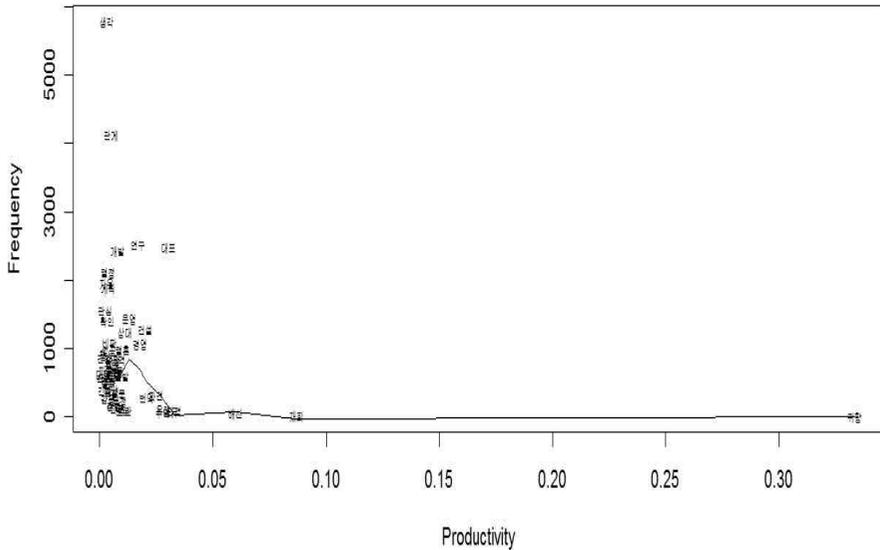


그림 8. 생산성 척도 P 와 빈도 (추세 곡선 포함)

반면에 구조 타입-토큰 비율과 빈도와 산포도의 추세 곡선을 살펴보면 왜곡되지 않는다. 그림 8과 같은 그래프를 구조 타입-토큰 비율과 빈도에 적용해보면 그림 9와 같다. 그림 9의 추세 곡선은 왜곡된 부분이 없으며, 그림 7과 같은 분포 곡선과 유사하다. 그림 8과 9를 종합해 보면, 생산성 척도 P 와 빈도의 추세 곡선은 왜곡된 구조를 보이는데 비해서 구조 타입-토큰 비율과 빈도의 증가 곡선은 왜곡되지 않는 것으로 보인다. 그러므로, 구조 타입-토큰 비율이 더 정확히 분포를 설명한다. 결과적으로 구조 타입-토큰 비율은 생산성 척도 P 와도 유사하면서도 실제 분포도 정확히 나타내기 때문에 생산성 척도 P 보다 더 유용한 기재이다.¹⁹

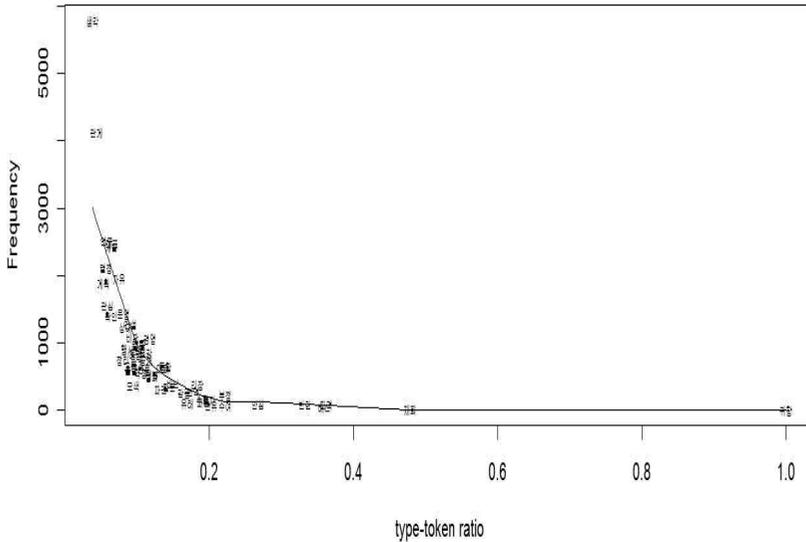


그림 9. 구조 타입-토큰 비율과 빈도 (추세 곡선 포함)

Baayen (1992, 1993)의 전역 생산성 지수 P^* 가 생산성 척도 P 와 어휘 유형 V 의 두 개의 차원의 표현이므로, 이는 그림 6, 8과 같은 형태의 그래프가 된다. 이 유형의 그래프에서 생산성 척도 P 와 빈도는 왜곡된 측면이 포함되어 있는데 비해서, 구조 타입-토큰 비율과 빈도는 분포를 설명하기에 적절하다. 따라서 생산성 지수 P^* 의 공간을 구조 타입-토큰으로 대체하는 것의 설명력이 크다고 할 수 있다.

지금까지 서술어/비서술어 명사 구조의 생산성을 살펴보았는데, 생산성은 서술어/비서술어 구분과 연관성이 없다. 또한 어휘 자질인 빈도, 다의성 목록, 명사 포함 유무, hapax legomena 등과도 연관성이 없다. 어휘의 의미-통사 구조를 나타내는 특징을 수치화한 타입-토큰 비율과 연관성이 있다. Baayen (1992, 1993)의 경우에 생산성 척도 P 와 어휘 유형의 두 개의 차원인 전역 생산성 지수 P^* 평면에서 설정된 일정한 조건으로 생산성을 결정한다. 여기서 생산성 척도 P 대신에 타입-토큰 비율을 활용하면 그림 10과 같다. 만약 기준이 타입-토큰 비율이 0.04이상, 어휘 유형 빈도가 300이상으로 설정하고, 기준을 통과한 어휘의 경우에 생산성이 있다고 설정하면 그림 10과 같다. 그림 10에서는 강범모(2000)에서 제시한 서술어/비서술어 목록을 합쳐서 그래프로 나타내었다.

19 이외에 강범모(2000)의 서술어/비서술어 명사, 이선웅(2005)의 비서술어 명사에도 같은 내용을 적용했는데, 동일한 결과가 산출되었다.

결되었다. 이 연구는 자료에 근거한 연구로 경험적 논점을 취하였다. 앞서 밝힌 바와 같이 서술어/비서술어 구분은 언어에 대한 이해에서 중요하다. 한국어에 대한 이전의 연구가 선형적 연구 기반으로 서술어/비서술어 구분에 대한 주장을 중심으로 하였다. 이에 필자는 앞으로의 연구에서 경험적 기반으로 서술어/비서술어 구분에 대한 연구를 다루고자 한다.

참고문헌

- 강범모, 차준경. 1996. “코퍼스, 파생어, 생산성”. 『인지과학』 7(2): 21-37.
- 강범모. 2000. “서술 명사의 기준과 의미 구조”. 2000년 한국언어정보학회 여름학술대회 발표논문집, 71-83.
- 강범모. 2001. “술어 명사의 의미 구조”. 『언어학』 31: 3-29.
- 박현아, 강범모. 2006. “한국어 서술성 명사의 실현 양상”. 제18회 한글 및 한국어 정보처리 학술대회 논문집, 117-124.
- 최형용. 2000. “‘-적’ 파생어의 의미와 ‘-적’의 생산성”. 『형태론』 2(2): 215-237.
- 시정근. 1999. “규칙은 과연 필요 없는가?”. 『형태론』 1(2): 261-283.
- 시정근. 2006. “국어 형태론에서의 ‘생산성’ 문제에 대한 연구”. 『형태론』 8(2): 257-276.
- 이광호. 2006. “파생접미사의 생산성과 파생어 집합의 빈도특성”. 『어문연구』 34(3): 219-250.
- 이병규. 2009. 『한국어 술어 명사문 연구』. 한국문화사.
- 이선웅. 2005. 『국어 명사의 논항구조 연구』. 월인.
- 홍재성. 1999. “기능동사 구문 연구의 한 시각: 어휘적 접근”. 『인문논집』, 135-173.
- Anderson, M. 1982. Where is morphology? *Linguistic Inquiry* 13: 571-612.
- Aronoff, M. 1976. *Word Formation in Generative Grammar*. MIT Press.
- Baayen, H. and R. Liber. 1991. Productivity and English Derivation. *Linguistics* 29: 801-843.
- Barddal, J. 2008. *Productivity: Evidence from Case and Argument Structure in Icelandic*. John Benjamins.
- Baayen, H. 1992. Quantitative aspects of morphological productivity. *Yearbook of Morphology* 1991: 109-149.
- Baayen, H. 1993. On frequency, transparency, and productivity. *Yearbook of Morphology* 1992: 181-208.
- Baayen, H. 2001. *Word Frequency Distributions*. Kluwer Publishers.
- Baroni, M. and S. Evert. 2006. Counting words. Paper presented in *European Summer School in Logic, Language and Information 2006*. (Also available at

- <http://zipfr.r-forge.r-project.org>).
- Bresnan, J., Cueni, A., Nikitina, T., and R. H. Baayen. 2007. Predicting the dative alternation. Bouma, G. and Kraemer, I. and Zwarts, J. (Eds.), *Cognitive Foundations of Interpretation, Royal Netherlands Academy of Arts and Sciences*, 69-94.
- Chomsky, N. 1965. *Aspects of the Theory of Syntax*. MIT Press.
- Conwell, E. and K. Demuth. 2007. Early syntactic productivity: Evidence from dative shift. *Cognition* 103: 163-179.
- Davidson, D. 1980. *Essays on Actions and Events*. Oxford University Press.
- Di Sciullo and E. Williams. 1987. *On the Definition of Word*. MIT Press.
- Fillmore, C. 1977. The case for case reopened. In R. Cole (Ed.), *Current Issues in Linguistic Theory*, 76-138.
- Gries, S. and A. Stefanowitsch. 2003. Collostructions. *International Journal of Corpus Linguistics* 8(2): 209-243.
- Gries, S. and A. Stefanowitsch. 2004. Extending collostructional analysis. *International Journal of Corpus Linguistics* 9(1): 97-129.
- Jackendoff, R. 1990. *Semantic Structures*. MIT Press.
- Katz, J. J. 1972. *Semantic Theory*. Crowell.
- Levin, B. 1993. *English Verb Classes and Alternations*. Chicago University Press.
- Levin, R. and M. R. Rappaport. 2005. *Argument Realization*. Cambridge University Press.
- Minsky, M. 1974. A framework for representing knowledge. In P. Winston (ed.), *The Psychology of Computer Vision*. McGraw-Hill.
- Parsons, T. 1990. *Events in the Semantics of English*. MIT Press.
- Pustejovsky, J. 1995. *Generative Lexicon*. MIT Press.
- Wurzel, W. 1984. *Inflectional Morphology and Naturalness*. Reidel.

참고 자료

- 세종 의미 분석 말뭉치. <http://www.sejong.or.kr>.
- 표준 국어 대사전. <http://stdweb2.korean.go.kr/main.jsp>.
- Alice in Wonderland (L. Carroll). <http://www.gutenberg.org>.

김동성

(136-701) 서울시 성북구 안암동 145

고려대학교 언어정보연구소

E-mail: dsk202@korea.ac.kr

접수일자 : 2012. 06. 29

수정일자 : 2012. 08. 17

게재일자 : 2012. 08. 20